Chromosome-wise Protein Interaction Patterns and Their Impact on Functional Implications of Large-Scale Genomic Aberrations

Kirk, Isa Kristina; Weinhold, Nils; Belling, Kirstine González-Izarzugaza; Skakkebæk, Niels Erik; Jensen, Thomas Skøt; Leffers, Henrik; Juul, Anders ; Brunak, Søren

Published in:
Cell Systems

Link to article, DOI:
10.1016/j.cels.2017.01.001

Publication date:
2017

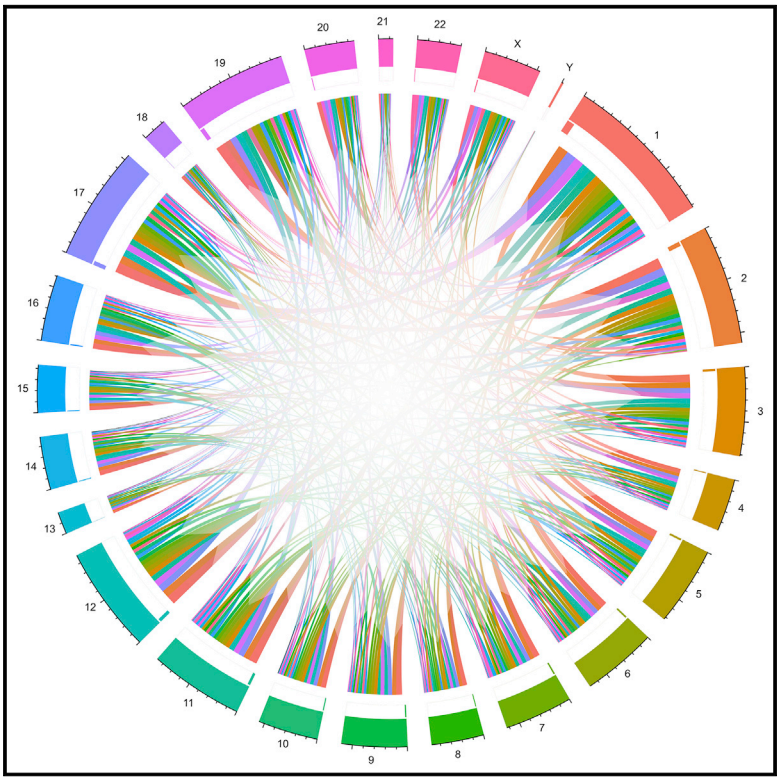
Document Version
Publisher's PDF, also known as Version of record

Cell Systems

Report

Chromosome-wise Protein Interaction Patterns and Their Impact on Functional Implications of Large-Scale Genomic Aberrations

Graphical Abstract



Highlights

- Chromosome-wise connectivity of protein interactions differs in the human genome
- The interactome may limit the impact of certain large-scale genomic aberrations
- New method to identify genes associated with trisomy 21 and other genomic aberrations

Authors

Isa Kristina Kirk, Nils Weinhold, Kirstine Belling, ..., Henrik Leffers, Anders Juul, Søren Brunak

Correspondence

soren.brunak@cpr.ku.dk

In Brief

Using the human interactome, we show that the variation in the chromosome-wise protein connectivity may limit the functional implications related to certain large-scale genomic aberrations, such as those seen in live born trisomies. This could likewise be the case for arm-level amplification in cancer. The study investigates trisomy 21 in depth, where the results demonstrate that protein interactions occur more rarely between proteins encoded by genes on chromosome 21 and, consequently, the third copy of chromosome 21 has a reduced functional impact on patients with trisomy 21.

Kirk et al., 2017, Cell Systems *4*, 357–364
March 22, 2017
http://dx.doi.org/10.1016/j.cels.2017.01.001

Cell Systems
Report

Chromosome-wise Protein Interaction Patterns and Their Impact on Functional Implications of Large-Scale Genomic Aberrations

Isa Kristina Kirk,[1,3,5] Nils Weinhold,[1,2,5] Kirstine Belling,[1,3,5] Niels Erik Skakkebæk,[4] Thomas Skøt Jensen,[1] Henrik Leffers,[4] Anders Juul,[4] and Søren Brunak[1,3,6,*]
[1]Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark, 2800 Lyngby, Denmark
[2]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
[4]Department of Growth and Reproduction, Rigshospitalet and University of Copenhagen, 2100 Copenhagen, Denmark
[5]Co-first author
[6]Lead Contact
*Correspondence: soren.brunak@cpr.ku.dk
http://dx.doi.org/10.1016/j.cels.2017.01.001

SUMMARY

Gene copy-number changes influence phenotypes through gene-dosage alteration and subsequent changes of protein complex stoichiometry. Human trisomies where gene copy numbers are increased uniformly over entire chromosomes provide generic cases for studying these relationships. In most trisomies, gene and protein level alterations have fatal consequences. We used genome-wide protein-protein interaction data to identify chromosome-specific patterns of protein interactions. We found that some chromosomes encode proteins that interact infrequently with each other, chromosome 21 in particular. We combined the protein interaction data with transcriptome data from human brain tissue to investigate how this pattern of global interactions may affect cellular function. We identified highly connected proteins that also had coordinated gene expression. These proteins were associated with important neurological functions affecting the characteristic phenotypes for Down syndrome and have previously been validated in mouse knockout experiments. Our approach is general and applicable to other gene-dosage changes, such as arm-level amplifications in cancer.

INTRODUCTION

Most proteins carry out their functions through physical interactions with other molecular components (Barabasi and Oltvai, 2004). Technological developments and data availability increasingly allow interactome-wide studies of disease mechanisms. In recent years, multiple studies have focused on the protein interactome that underlies diseases and their co-occurrences (Lage et al., 2007; Goh et al., 2007; Zhou et al., 2014; Wang et al., 2012; Menche et al., 2015). It is now the general understanding that many diseases are typically not caused by the activity of a single defect component, but that the altered component can affect a cascade of interacting components (Barabasi et al., 2011). Thus, it is of importance to study the general topology of the protein interactome to understand disease pathophysiology.

Gene expression alterations of individual members in protein networks cause stoichiometric imbalance and disrupt functionality (Birchler and Veitia, 2010). Recent advances of mass spectrometry technology allow for studying the quantitative protein interactome as a function of the underlying human proteome (Bantscheff et al., 2012; Wisniewski et al., 2014; Hein et al., 2015). A wide spectrum of interaction stoichiometries has been shown to associate with different levels of protein network stabilities. Stable network parts such as protein complexes have the most balanced stoichiometries, while more transient interactions tend to be associated with substoichiometric proportions (Hein et al., 2015). These two kinds of interaction stoichiometries play different roles in terms of global network topological properties and their impact on modularity and resilience (Hein et al., 2015). In particular, removal of substoichiometric interactions from the global network of protein interactions leads to rapid network fragmentation, whereas removal of 50% of the strongest edges causes hardly any network fragmentation (Hein et al., 2015).

Numerical abnormalities of chromosomes have drastic effects on gene dosages. In the case of additional human autosomes many cannot be suppressed sufficiently by compensatory regulation (Nagaoka et al., 2012). The phenotypic impact of genomic imbalances is hard to estimate using metrics such as chromosome length and gene count alone. For example, human X chromosome aneuploidies have high survival rates whereas chromosome 22 trisomies are very rarely observed in liveborns (Heinrich et al., 2013). Thus, it is of particular interest to show how aberrations of whole chromosomes or chromosome arms via gene-dosage alterations affect protein complex stoichiometry.

Here we used an annotated, highly accurate resource of experimentally derived protein-protein interaction (PPI) data (Li et al., 2017) to analyze how pairs of human chromosomes connect through protein interactions. We used the pairwise PPIs to define highly connected proteins with at least five

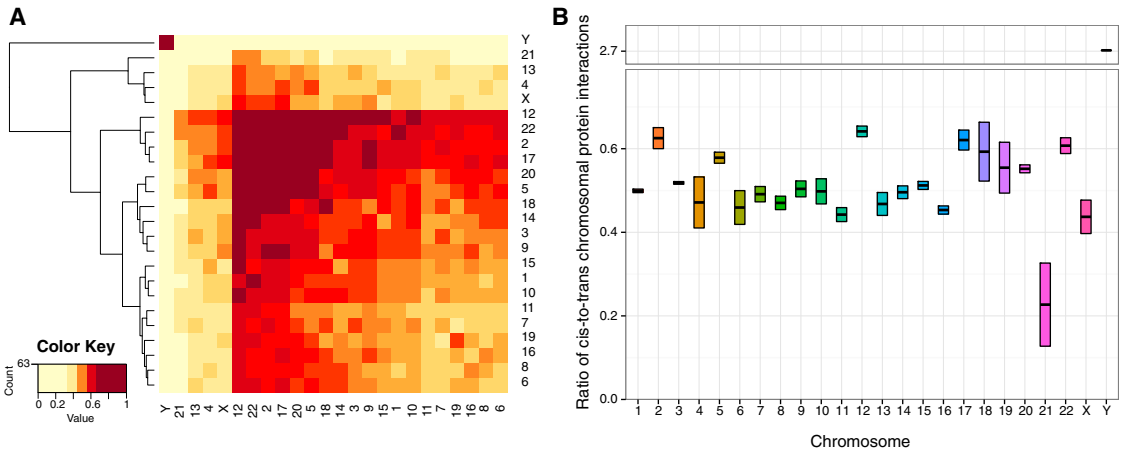


Figure 1. Patterns of Protein Interaction across Human Chromosomes

(A) Protein interactions between proteins encoded by different chromosomes. The heatmap visualizes pairwise similarities between all human chromosomes. Chromosomes were clustered based on normalized counts of *cis*- and *trans*-chromosomal protein interactions. Colors represent relative frequencies of protein interactions between pairs of chromosomes, ranging from light yellow (low) to dark red (high). Two main groups of chromosomes were identified using hierarchical clustering. Chromosome 21 was part of a small cluster of chromosomes with few *cis*- and *trans*-chromosomal interactions (together with chromosomes X, 4, and 13). Chromosome 21 had particularly few *cis*-chromosomal interactions compared with other chromosomes.

(B) Robustness analysis of chromosome-wise interactions using three protein interactome sets with different confidence score thresholds. The ratio of normalized *cis*- to *trans*-chromosomal protein interactions (connectivity ratio) was calculated for each chromosome using the three different subsets of protein interactions. Connectivity ratios for the three subsets are represented as horizontal bars for each chromosome.

interacting proteins as hubs (d $\geq$ 5). When used in combination with chromosomal information, hubs represent a framework for understanding how copy-number changes affect cellular regulation and execute particular cellular functions, both at the interaction and the global level. Since transcriptomics data reflects changes in copy numbers they can be used to study changes in protein abundance that potentially change the stoichiometry.

RESULTS

Proteins Encoded by Chromosome 21 Interact Infrequently with Each Other

We investigated the human interactome by using a protein interaction dataset, InWeb_IM (Li et al., 2017) containing verified and benchmarked PPIs. We calculated the number of observed PPIs between pairs of proteins encoded by the same chromosome (termed *cis*-chromosomal interactions), as well as the number of PPIs between proteins encoded by different chromosomes (termed *trans*-chromosomal interactions). Since chromosome size varies greatly with respect to numbers of genes (Figure S1A), we normalized observed counts of *cis*- and *trans*-chromosomal interactions by the number of theoretically possible interactions (Figure S1B and STAR Methods).

When comparing chromosomes based on normalized counts of *cis*- and *trans*-chromosomal interactions (Figure 1A), we observed that a group of chromosomes (4, 13, 21, and X) had fewer PPIs compared with all other chromosomes. These chromosomes were characterized by few *cis*-chromosomal and few *trans*-chromosomal interactions where chromosome 21 had the lowest number of interactions overall (14 *cis*- and 5,207 *trans*-chromosomal interactions). Notably, three out of the four chromosomes with reduced numbers of interactions lead to liveborn trisomies (13, 21, and X). It is well known that additional copies of chromosomes 21 and X lead to morphological and phenotypic changes, but are tolerated and generally compatible with life, whereas liveborn trisomies of chromosome 13 and 18 only survive for about a week (Brewer et al., 2002). Chromosome Y has the fewest protein-coding genes (N = 46) that mostly encode products specific for male sex determination and male germ cell production (Quintana-Murci and Fellous, 2001). Genes located on chromosome Y primarily interact with each other, reflecting their close functional relationship (Figure 1A). Aneuploidies involving chromosome Y are compatible with long-term survival as well.

To test the robustness of the results shown in Figure 1A, we compared all human chromosomes based on *cis*- and *trans*-chromosomal interactions. A connectivity ratio was defined as the ratio of normalized *cis*-chromosomal interactions to normalized *trans*-chromosomal interactions and calculated for each chromosome using three different subsets of the InWeb_IM interactome (Figure 1B and STAR Methods). The subsets were selected based on different thresholds of PPI confidence scores. The vast majority of chromosomes had roughly as many *cis*-chromosomal as *trans*-chromosomal interactions (mean = 1.02, SD = 0.18 when excluding chromosome Y). Notably, chromosome 21 had the lowest connectivity ratio among all chromosomes (>3 SD below average), and had a significantly lower ratio than expected (p = 0.012). In an independent randomization analysis chromosome 21 was the only chromosome with significantly lower connectivity ratio (p = 0.04). These observations suggest that the structure of the interactome suppresses the impact of the additional chromosome in trisomy 21, which is characteristic for Down syndrome (DS), more than it does in other trisomies. Consequently, long-term survival of patients with trisomy 21 might be promoted by the extremely low ratio of

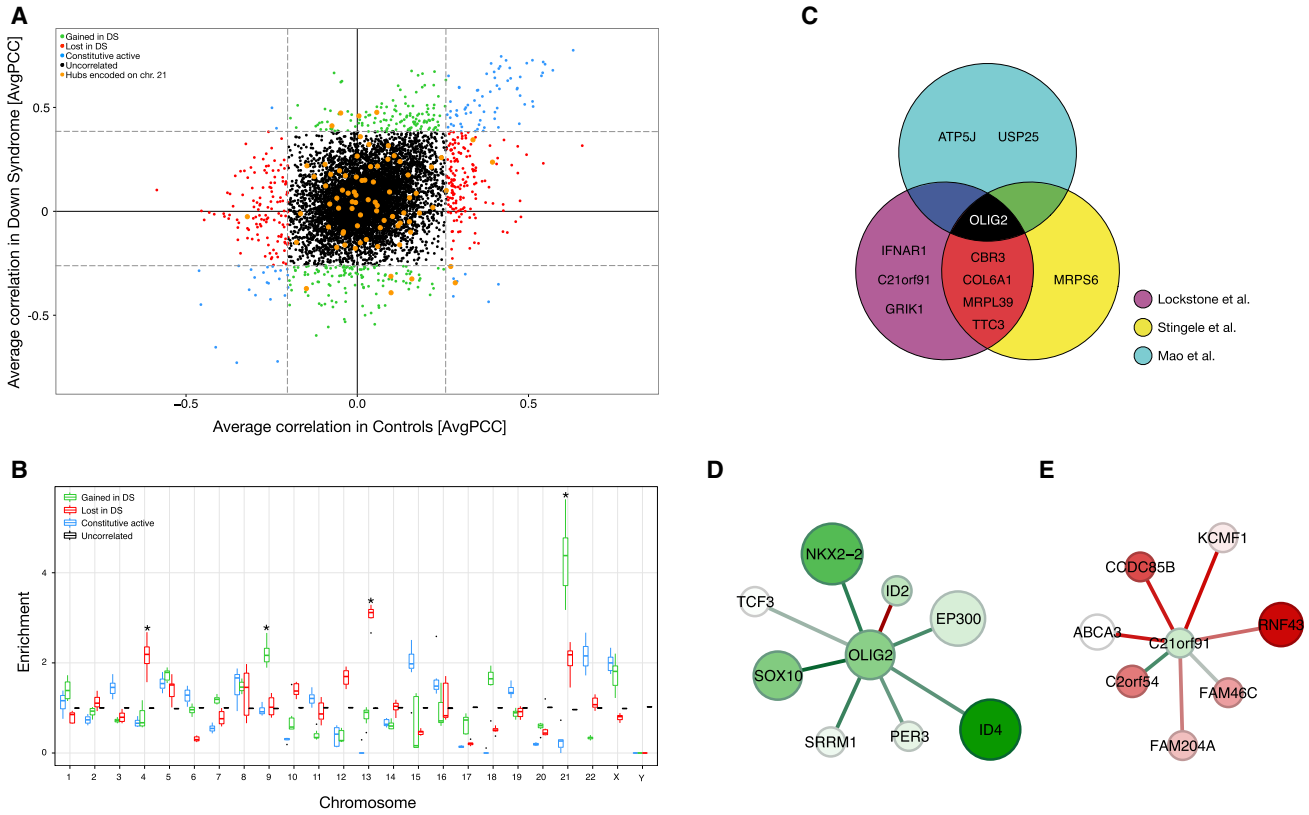


Figure 2. Activity of Proteins in Down Syndrome

(A) Correlation of gene expression between a hub protein and its interactors. Points represent protein hubs according to the average correlation (*AvgPCC*) calculated for DS patients (y axis) and controls (x axis). Correlated hubs (the 5% most correlated with an *AvgPCC* > 2 SD from mean) were considered active and were divided into three groups: gained in DS patients (green), lost in DS (red), and constitutively active (blue). Hubs encoded by genes on chromosome 21 are marked in orange. Hubs with *AvgPCC* not significantly different (<2 SD) from the mean in DS patients as well as controls were considered not active in brain tissue.

(B) Chromosome-wise enrichment analysis of active hubs. The asterisks (*) mark the chromosomes significantly enriched in the groups: gained in DS (chromosome 9 and 21) and lost in DS (chromosome 4 and 13).

(C) Hubs gained in DS identified in two or more gene expression datasets.

(D and E) Hubs OLIG2 and c21orf91 and interacting proteins. Node colors represent logFC between DS patients and controls (green, higher expression in DS; red, higher expression in controls). Larger nodes indicate whether there is a significant difference in logFC between DS patients and controls. Edge colors represent the pairwise correlation between hub and interactor in DS patients (green, positive correlation; red, negative correlation; gray, no correlation).

cis- to *trans*-chromosomal interactions. The low ratio could prevent over-abundant proteins encoded by chromosome 21 from forming interactions with equally abundant proteins, which would significantly disturb the balance of cellular protein complex abundances.

We checked for possible confounding factors influencing the connectivity ratio: (1) increased frequency of PPIs between paralogs (Figure S2A); (2) a possible correlation between topological-associated domains and PPIs (Figure S2B); and (3) a possible chromosomal bias in the calculation of the connectivity ratio (Figures S2C and S2D). None of the three factors were found to influence the chromosomal connectivity ratio.

Trisomy 21 Modulates Protein Activity in Adult Brain Tissue

To understand how the transcriptional changes caused by trisomy 21 affect the protein network structure, we implemented a network analysis method to identify functional hub proteins, proteins connected by at least five interaction partners, in DS patients (Han et al., 2004). Since most proteins function in concert with other proteins, we identified active hub proteins by including gene expression information for both the hub and its interaction partners. Expression correlation between the hub and its interaction partners was used as a means to deem the hub being functionally active. This method has previously been used to identify therapeutically targetable proteins in breast cancer (Taylor et al., 2009), and to prioritize disease genes across various tissues (Bornigen et al., 2013). Using published expression data from brain tissue of both adult DS patients and controls, we calculated the average Pearson correlation coefficient (*AvgPCC*) in DS patients and controls separately for each hub (STAR Methods). *Z* scores were obtained by standardizing *AvgPCC* values, and the hub proteins were ranked based on their absolute *Z* score in DS patients and controls, respectively.

By combining ranked hub proteins in DS patients and controls (Figure 2A), we identified three groups of hubs that were: (1) functionally *gained* in DS patients, i.e., highly correlated in DS patients but not in controls (267 hubs); (2) functionally *lost* in

Table 1. Correlation-Based Classification Criteria for Groups of Proteins

Group	Direction of Correlation in DS[a]	Direction of Correlation in Controls[a]	No.	Protein Activity
Gained in DS	positive	negative	267	active in DS, inactive in controls
Lost in DS	negative	positive	287	inactive in DS, active in controls
Constitutively active	positive	positive	100	active in DS and controls
Uncorrelated	–	–	6,727	ND

[a]Deviating > 2 SD from mean (the 5% most correlated hubs). ND, not determined.

DS patients, i.e., highly correlated in controls, but not in DS patients (287 hubs); (3) *constitutively* active (N = 100), i.e., highly correlated in both DS patients and controls. Hubs in these categories deviated significantly (>2 SD, equal to the 5% most correlated) from the mean and were considered functionally active. Hubs that differed by less than 2 SD from the mean in both DS patients and controls were considered uncorrelated, and therefore functionally inactive in both states (Table 1). Each of the groups mentioned above contained a small number of hubs encoded on chromosome 21 (Figure 2A). In contrast to single-gene expression changes (Figure S3A), a relatively small number of hubs encoded by chromosome 21 showed significant changes in activity, indicating that they contribute to a gained or lost function in DS. Further, when testing whether particular chromosomes were enriched for hubs either gained or lost in DS patients, we found that hubs gained in DS were enriched on chromosome 21 ($p = 0.03$) and 9 ($p = 0.04$), whereas hubs lost in DS were enriched on chromosome 4 ($p = 0.003$) and 13 ($p = 0.007$) (Figure 2B and Table S1).

Highly Correlated Hub Proteins Encoded on Chromosome 21 Were Associated with Neural Diseases and Development

Using the absolute *Z* scores for the expression correlation, we ranked the hubs encoded by chromosome 21 (Table S1). In total, 11 of 78 hubs encoded by chromosome 21 were either gained (N = 8) or lost (N = 3) in DS patients. We applied the same approach to gene expression data from trisomic cell lines (Stingele et al., 2012) as well as fetal brain tissue (Mao et al., 2005), and observed that five correlated hub proteins encoded on chromosome 21 were identified as gained in DS by at least two studies (Figure 2C). The protein OLIG2 was identified in all three datasets and has previously been shown to cause developmental brain defects in DS and to contribute to the development of leukemia among other cancers (Lin et al., 2005). Importantly, high expression levels of OLIG2 limit the proliferation of neural progenitors and therefore advance a reduction of neural cells and brain size in DS patients (Lu et al., 2012). OLIG2 and its interactors were predominantly positively correlated (Figure 2D) and based on these observations it is reasonable to assume that the stoichiometric correlations of OLIG2 contribute to neural deficiencies in DS patients.

In addition to OLIG2, four other correlated hubs encoded by chromosome 21 were gained in DS and identified by analysis of two gene expression datasets: CBR3 is a carbonyl reductase relevant to Alzheimer's disease (Watanabe et al., 1998) and associated with heart defects in DS (Liu et al., 2014), COL6A1 has been linked to congenital heart defects and ocular anomalies in DS (Bromham et al., 2002; Davies et al., 1995), MRPL39 is located in the trisomic region of a popular mouse model for DS (Ramakrishna et al., 2009), and TTC3 is located in the so-called DS critical region and is an inhibitor of neural development (Berto et al., 2007). DYRK1A had equally many positive and negative correlating interactions as well as many uncorrelated interactions, and was therefore not in top 5% of the most correlated hub proteins in our data (*AvgPCC* < 2 SD from mean) (Figure S3B).

Notably, the most significantly correlated hub with gain-of-function characteristics (i.e., stronger correlation in DS patients) was c21orf91 (Figure 2E), an open reading frame only very recently characterized to be an important new player in neural development (Li et al., 2016). Previously it has been reported to be highly expressed in the brain tissue of a patient with partial tetrasomy of chromosome 21 (Rost et al., 2004) and has also been implicated in eye development (Godbout et al., 2003). Another recent paper (Olmos-Serrano et al., 2016) showed that this open reading frame was upregulated in postmortem brains from DS compared with controls and present in a co-expressed module regulating action potential and axon ensheathment.

We performed enrichment analyses for disease association for the chromosome 21 encoded hubs that were gained or lost in DS. No disease association was found for the three hubs lost in DS. However, for the hubs gained in DS, we found that TTC3, CBR3, IFNAR1, GRIK1, and COL6A were enriched in DS (adj. $p = 1.28 \times 10^{-8}$), OLIG2, TTC3, and COL6A were enriched in nervous systems diseases (adj. $p = 0.006$), and OLIG2 and GRIK1 were enriched in mental disorders (adj. $p = 0.049$). This supports that the top-ranked genes gained in DS are implicated in neurological mechanisms.

We compared the three transcriptomics datasets to mass-spectrometric measurements of protein abundances. Ideally, one would use quantitative proteomics measurements covering the entire proteome to accurately analyze protein network stoichiometries. Unfortunately, such data are yet not available for the relevant tissues in this study. A recent analysis of responses to aneuploidies in human cells (Stingele et al., 2012) generated proteomics data covering around 6,000 proteins, of which 53 proteins are encoded by chromosome 21. Of the five highly correlated hubs encoded by chromosome 21 and gained in DS, protein abundance measurements were available for COL6A1, CBR3, and MRPL39. All three proteins had log2 protein expression ratios (trisomy 21 versus diploid cell lines) above zero, indicating higher protein expression in trisomic cell lines compared with cell lines with normal karyotypes, thus suggesting important roles in DS patients.

Analysis of the proteomics data (Hein et al., 2015) did, however, not show differences between chromosome 21 and other human chromosomes with respect to the distribution of stoichiometric and substoichiometric protein interactions. Based on the limited coverage of chromosome 21 in the proteomics data, we did not observe any pattern in protein interaction stoichiometry

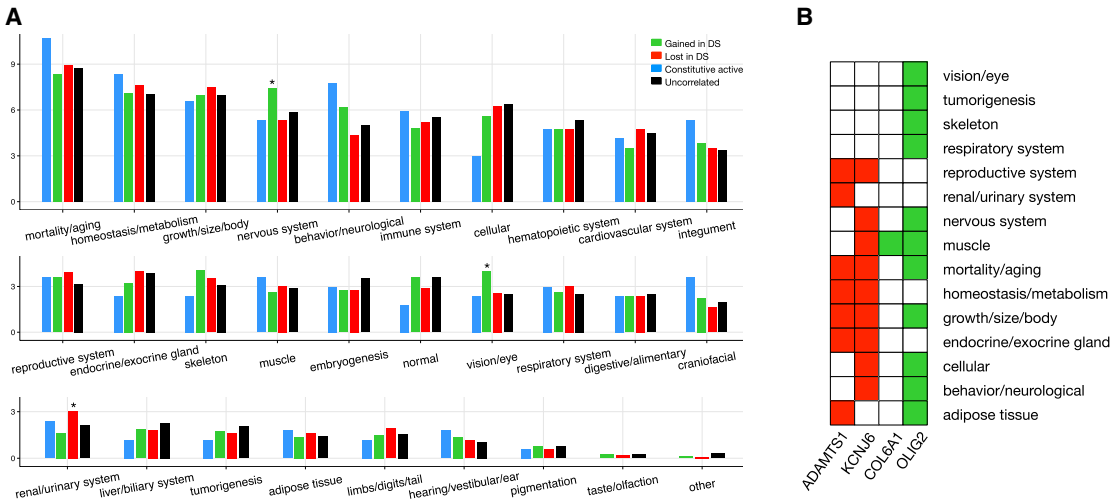


Figure 3. Distribution of Phenotypes Among Active Proteins

(A) Phenotypes were ordered based on enrichment for hubs encoded by genes on chromosome 21 in all four types: gained in DS, lost in DS, constitutively active, or uncorrelated. A hypergeometric test was performed to identify phenotypes significantly enriched for hubs encoded by genes on chromosome 21. The nervous system and vision/eye were significantly enriched for hubs gained in DS, while the renal/urinary system was enriched for hubs lost in DS. Significantly enriched categories are marked with an asterisk (*).

(B) Hubs that were gained or lost in DS and significantly associated phenotypes are highlighted in green (gained in DS) and red (lost in DS). Hubs encoded by genes on chromosome 21 that were not associated with any MGI phenotype are not shown.

that would explain the distinctive role of chromosome 21 other than the observations described above, i.e., generally low number of PPIs as well as a low number of *cis*-chromosomal interactions relative to *trans*-chromosomal interactions.

Functionally Altered Hub Proteins Affect DS Phenotypes

To further understand which phenotypes were affected most by the correlated hub proteins gained or lost in DS, we used gene knockout information from the Mouse Genome Informatics (MGI) database. We tested all phenotypes from the first level of the MGI ontology (29 phenotypes in total) for enrichment of the correlated hubs encoded by chromosome 21 using a hypergeometric test (Figure 3A). Two phenotypes were significantly enriched for proteins gained in DS: nervous system and vision/eye. This is in agreement with the fact that central nervous system (CNS) changes are fundamental to phenotypic implications of trisomy 21. DS patients are commonly affected by intellectual disability and a variety of other neurological abnormalities. These phenotypic alterations include, but are not limited to, changes in number, structure, and function of neurological cells (Morice, 2010). Notably, disturbances in the gastrointestinal system observed in DS patients are caused by disorders of the enteric nervous system and could therefore be linked to changes in the CNS, which we observed here. Esophageal motor dysfunction, chronic constipation, and Hirschsprung's disease are all observed in DS patients (Moore, 2008). The other enriched phenotype, vision/eye, is also supported by the literature on DS phenotypes, which describes significant vision deficits, including reduced contrast sensitivity, acuity, and ability to discriminate color (Krinsky-McHale et al., 2014). Moreover, children with DS have significantly lower acuity thresholds as well significantly reduced contrast sensitivity than controls (John et al., 2004). In contrast to the nervous system and vision/eye phenotypes, which were enriched for correlated hubs gained in DS, the renal/urinary system was enriched for hubs lost in DS. Some patients with DS are indeed affected by a variety of renal and urinary tract abnormalities including obstructive uropathy (Ariel et al., 1991) and fetal pyelectasis, a dilation of the renal pelvis, which have been observed in 25% of fetuses with DS (Benacerraf et al., 1990). Figure 3B summarizes significant hubs gained or lost in DS and their associated phenotypes.

DISCUSSION

Our results suggest that the human interactome has an intrinsic structure and that different chromosomes participate differently in it. By comparing PPIs on a chromosome-by-chromosome basis, we observed that proteins encoded by chromosome 21 have few *cis*-chromosomal interactions, and that the ratio of *cis*- to *trans*-chromosomal interactions was much lower for chromosome 21 than any other human chromosome. It appears that chromosome 21 is tightly embedded into the overall structure of the interactome, but fulfills a unique role of its own, which potentially limits the regulatory and phenotypic impact of trisomy 21 compared with other autosomal trisomies, and might also contribute to explaining the relatively high viability of patients with DS. We used a previously published method to combine human interactome data and gene expression data from adult brain tissue. This analysis identified highly correlated hub proteins (proteins with at least five interactors in the interactome) gained, lost, or constitutively active in DS compared with controls. Of particular interest was OLIG2, a hub gained in DS, which was validated in two additional gene expression datasets. OLIG2 is known to play an important role in neurologic development and was enriched in mental disorders and nervous system diseases. Notably, this protein was not identified in the original study based

on gene expression data alone (Lockstone et al., 2007). We therefore suggest that this method of analyzing protein hub activity is complementary to existing single-gene approaches, and may be applicable in other scenarios where gene expression data can be combined with highly accurate protein interaction data.

To computationally analyze the structure of the complete interactome connecting chromosomes one would ideally account for microRNA genes and other non-protein-coding genes as well. However, it is currently not possible to accurately determine detailed interaction patterns for these genes, and the analyses presented here are therefore based only on the known protein-based interactome. The number of interactions from public databases in the confidence-scored interactome correlates highly with the number of genes on each chromosome (Figure S1C). While there are known biases in large high-throughput proteomics screens, such as weak coverage of low-abundance proteins or proteins with very few or no tryptic peptides, these biases are unlikely to be chromosome specific. Therefore, the analyses presented here are based on a reasonably unbiased, high-confidence sampling of the human protein interactome.

Gene copy numbers vary considerably among human individuals (Handsaker et al., 2015), and this variation is highly relevant in many diseases. This study suggests that our understanding of phenotypic consequences of copy-number changes is limited by analysis of single genes and their pathways (Kim et al., 2015). It is equally important to account for protein abundance changes that can affect protein network stoichiometry, which can suppress or amplify the impact of aneuploidies. Since most gene products exert their function as part of multi-subunit protein complexes, copy-number changes of protein-coding genes are of particular interest. Balancing subunit abundances in protein complexes (stoichiometry) is required for assembly of functional complexes (Hein et al., 2015), and regulation of this balance is used to in-/activate and modify functional complexes, for example in the case of the cell cycle (Olsen et al., 2010). It was shown recently that proteins compete for binding to hub proteins at critical network branchpoints, and that differences in protein abundances lead to differences in signaling phenotypes (Kiel et al., 2013). Although we did not define hub proteins based on known protein complexes but on pairwise PPIs obtained from the interactome set used, we showed that combining these with expression analysis still enables us to determine hub proteins implicated in DS characteristics.

Aneuploidies such as trisomy 21 constitute a special case of copy-number variation since they add a whole chromosome to the karyotype. Model systems have been used extensively to study cellular imbalances caused by additional copies of whole chromosomes in yeast and in mice (Galdzicki et al., 2001). Many human cancers are characterized by copy-number changes including longer, arm-level amplifications as well as shorter, highly focal amplifications (Beroukhim et al., 2010). Analysis of somatic mutations from whole-genome sequences suggests that cancer genomes undergo stepwise transformation during which additional, evolutionary advantageous events are acquired successively (Landau et al., 2013). Gene-dosage changes caused by amplification or deletion of important cancer genes are one such event. Apart from cancer, several other genomic diseases such as Alzheimer's disease and schizophrenia are characterized by copy-number alterations (Cook and Scherer, 2008). More recently, large copy-number variants were shown to compound each other and lead to a more severe clinical presentation (Girirajan et al., 2012). One major challenge is to identify the causative genes in large copy-number variants in cancer as well as in neurocognitive disorders.

Our analyses evaluate the possibility of a built-in robustness that is inherent to the structure of the human protein interactome. This type of analysis is applicable in any study where gene-dosage changes are relevant, such as arm-level aneuploidies and even more focal copy-number changes, which are both common in a wide range of cancers. Currently, strategies for DS management and therapy are directed toward identification of alterations at both chromosomal and molecular levels (de la Torre et al., 2016; Delabar et al., 2016), highlighting the relevance of approaches for identifying potential therapeutic targets described in this study. A recent study presented a dosage compensation strategy in DS pluripotent stem cells using XIST-based silencing targeting the additional chromosome (Cook and Scherer, 2008). This strategy was characterized as the major first step toward "chromosome therapy," for which the present paper also provides additional insight.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Protein Interaction Data
  - Chromosome-Wise Connectivity Ratio
  - Global Similarity of Chromosomal PPI Profiles
  - Highly Connected Hub Proteins
  - Gene Expression Data
  - Expression Correlation of Hub Proteins
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Significance of the Chromosome-wise Connectivity Ratio (CR)
  - Disease Enrichment in Chromosome 21 Encoded Hubs
  - Gene Enrichment
  - Functional Sub-Classification of Correlated Hub Proteins
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, one table, and supplemental text and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2017.01.001.

AUTHOR CONTRIBUTIONS

N.W., A.J., N.E.S., and S.B. conceived the study; N.W., I.K.K., K.B., and S.B. designed the experiments; I.K.K. and N.W. implemented the methods; N.W., I.K.K., K.B., and S.B. analyzed the data; N.W., I.K.K., K.B., and S.B. wrote the manuscript; N.E.S., T.S.J., H.L., and A.J. provided scientific and conceptual advice. All authors reviewed and approved the manuscript. The authors declare no competing financial interests.

ACKNOWLEDGMENTS

We would like to acknowledge funding from the FP7 grant SyBoSS (EU seventh Framework G.A. no. 242129) as well as the Novo Nordisk Foundation (grant agreement NNF14CC0001). Further, we would like to acknowledge Matthias Mann for facilitating data access and Lars Juhl Jensen for useful comments.

Received: May 7, 2016
Revised: October 23, 2016
Accepted: January 5, 2017
Published: February 15, 2017

REFERENCES

Ariel, I., Wells, T.R., Landing, B.H., and Singer, D.B. (1991). The urinary system in Down syndrome: a study of 124 autopsy cases. Pediatr. Pathol. *11*, 879–888.

Bantscheff, M., Lemeer, S., Savitski, M.M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal. Bioanal. Chem. *404*, 939–965.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Barabasi, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat. Rev. Genet. *12*, 56–68.

Benacerraf, B.R., Mandell, J., Estroff, J.A., Harlow, B.L., and Frigoletto, F.D., Jr. (1990). Fetal pyelectasis: a possible association with Down syndrome. Obstet. Gynecol. *76*, 58–60.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Berto, G., Camera, P., Fusco, C., Imarisio, S., Ambrogio, C., Chiarle, R., Silengo, L., and Di Cunto, F. (2007). The Down syndrome critical region protein TTC3 inhibits neuronal differentiation via RhoA and Citron kinase. J. Cell Sci. *120*, 1859–1867.

Birchler, J.A., and Veitia, R.A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol. *186*, 54–62.

Bornigen, D., Pers, T.H., Thorrez, L., Huttenhower, C., Moreau, Y., and Brunak, S. (2013). Concordance of gene expression in human protein complexes reveals tissue specificity and pathology. Nucleic Acids Res. *41*, e171.

Brewer, C.M., Holloway, S.H., Stone, D.H., Carothers, A.D., and Fitzpatrick, D.R. (2002). Survival in trisomy 13 and trisomy 18 cases ascertained from population based registers. J. Med. Genet. *39*, e54.

Bromham, N.R., Woodhouse, J.M., Cregg, M., Webb, E., and Fraser, W.I. (2002). Heart defects and ocular anomalies in children with Down's syndrome. Br. J. Ophthalmol. *86*, 1367–1368.

Cook, E.H., Jr., and Scherer, S.W. (2008). Copy-number variations associated with neuropsychiatric conditions. Nature *455*, 919–923.

Davies, G.E., Howard, C.M., Farrer, M.J., Coleman, M.M., Bennett, L.B., Cullen, L.M., Wyse, R.K., Burn, J., Williamson, R., and Kessling, A.M. (1995). Genetic variation in the COL6A1 region is associated with congenital heart defects in trisomy 21 (Down's syndrome). Ann. Hum. Genet. *59*, 253–269.

de la Torre, R., De Sola, S., Hernandez, G., Farre, M., Pujol, J., Rodriguez, J., Espadaler, J.M., Langohr, K., Cuenca-Royo, A., Principe, A., et al. (2016). Safety and efficacy of cognitive training plus epigallocatechin-3-gallate in young adults with Down's syndrome (TESDAD): a double-blind, randomised, placebo-controlled, phase 2 trial. Lancet Neurol. *15*, 801–810.

Delabar, J.M., Allinquant, B., Bianchi, D., Blumenthal, T., Dekker, A., Edgin, J., O'Bryan, J., Dierssen, M., Potier, M.C., Wiseman, F., et al. (2016). Changing paradigms in down syndrome: the first international conference of the trisomy 21 research society. Mol. Syndromol. *7*, 251–261.

Galdzicki, Z., Siarey, R., Pearce, R., Stoll, J., and Rapoport, S.I. (2001). On the cause of mental retardation in Down syndrome: extrapolation from full and segmental trisomy 16 mouse models. Brain Res. Brain Res. Rev. *35*, 115–145.

Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S., et al. (2012). Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N. Engl. J. Med. *367*, 1321–1331.

Godbout, R., Andison, R., Katyal, S., and Bisgrove, D.A. (2003). Isolation of a novel cDNA enriched in the undifferentiated chick retina and lens. Dev. Dyn. *227*, 409–415.

Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabasi, A.L. (2007). The human disease network. Proc. Natl. Acad. Sci. USA *104*, 8685–8690.

Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P., and Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88–93.

Handsaker, R.E., Van Doren, V., Berman, J.R., Genovese, G., Kashin, S., Boettger, L.M., and McCarroll, S.A. (2015). Large multiallelic copy number variations in humans. Nat. Genet. *47*, 296–303.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.

Heinrich, T., Nanda, I., Rehn, M., Zollner, U., Frieauff, E., Wirbelauer, J., Grimm, T., and Schmid, M. (2013). Live-born trisomy 22: patient report and review. Mol. Syndromol. *3*, 262–269.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Jr., Wright, M.W., et al. (2004). Gene map of the extended human MHC. Nat. Rev. Genet. *5*, 889–899.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. *31*, e15.

John, F.M., Bromham, N.R., Woodhouse, J.M., and Candy, T.R. (2004). Spatial vision deficits in infants and children with Down syndrome. Invest. Ophthalmol. Vis. Sci. *45*, 1566–1572.

Kiel, C., Verschueren, E., Yang, J.S., and Serrano, L. (2013). Integration of protein abundance and structure data reveals competition in the ErbB signaling network. Sci. Signal. *6*, ra109.

Kim, S., Kon, M., and Kang, H. (2015). A method for generating new datasets based on copy number for cancer analysis. Biomed. Res. Int. *2015*, 467514.

Krinsky-McHale, S.J., Silverman, W., Gordon, J., Devenny, D.A., Oley, N., and Abramov, I. (2014). Vision deficits in adults with Down syndrome. J. Appl. Res. Intellect. Disabil. *27*, 247–263.

Lage, K., Karlberg, E.O., Storling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tumer, Z., Pociot, F., Tommerup, N., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. *25*, 309–316.

Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell *152*, 714–726.

Li, S.S., Qu, Z., Haas, M., Ngo, L., Heo, Y.J., Kang, H.J., Britto, J.M., Cullen, H.D., Vanyai, H.K., Tan, S.S., et al. (2016). The HSA21 gene EURL/C21ORF91 controls neurogenesis within the cerebral cortex and is implicated in the pathogenesis of Down syndrome. Sci. Rep. *6*, 29514.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Staerfeldt, H.H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods *14*, 61–64.

Lin, Y.-W., Deveney, R., Barbara, M., Iscove, N.N., Nimer, S.D., Slape, C., and Aplan, P.D. (2005). OLIG2 (BHLHB1), a bHLH transcription factor, contributes to leukemogenesis in concert with LMO1. Cancer Res. *65*, 7151–7158.

Liu, C., Morishima, M., Jiang, X., Yu, T., Meng, K., Ray, D., Pao, A., Ye, P., Parmacek, M.S., and Yu, Y.E. (2014). Engineered chromosome-based genetic mapping establishes a 3.7 Mb critical genomic region for Down syndrome-associated heart defects in mice. Hum. Genet. *133*, 743–753.

Lockstone, H.E., Harris, L.W., Swatton, J.E., Wayland, M.T., Holland, A.J., and Bahn, S. (2007). Gene expression profiling in the adult Down syndrome brain. Genomics *90*, 647–660.

Lu, J., Lian, G., Zhou, H., Esposito, G., Steardo, L., Delli-Bovi, L.C., Hecht, J.L., Lu, Q.R., and Sheen, V. (2012). OLIG2 over-expression impairs proliferation of human Down syndrome neural progenitors. Hum. Mol. Genet. *21*, 2330–2340.

Mao, R., Wang, X., Spitznagel, E.L., Frelin, L.P., Ting, J.C., Ding, H., Kim, J.-W., Ruczinski, I., Downey, T.J., and Pevsner, J. (2005). Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. Genome Biol. *6*, R107.

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., and Barabasi, A.L. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science *347*, 1257601.

Moore, S.W. (2008). Down syndrome and the enteric nervous system. Pediatr. Surg. Int. *24*, 873–883.

Morice, E. (2010). Relevance of animal models in the study of human pathologies: a mouse model of Down syndromes. Biol. Aujourdhui *204*, 3–8.

Nagaoka, S.I., Hassold, T.J., and Hunt, P.A. (2012). Human aneuploidy: mechanisms and new insights into an age-old problem. Nat. Rev. Genet. *13*, 493–504.

Olmos-Serrano, J.L., Kang, H.J., Tyler, W.A., Silbereis, J.C., Cheng, F., Zhu, Y., Pletikos, M., Jankovic-Rapan, L., Cramer, N.P., Galdzicki, Z., et al. (2016). Down syndrome developmental brain transcriptome reveals defective oligodendrocyte differentiation and myelination. Neuron *89*, 1208–1222.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. *3*, ra3.

Quintana-Murci, L., and Fellous, M. (2001). The human Y chromosome: the biological role of a "functional Wasteland". J. Biomed. Biotechnol. *1*, 18–24.

Ramakrishna, N., Meeker, H.C., Li, S., Brown, W.T., Rao, R., and El Idrissi, A. (2009). Upregulation of beta-catenin expression in Down syndrome model Ts65Dn mouse brain. Neuroscience *161*, 451–458.

Rost, I., Fiegler, H., Fauth, C., Carr, P., Bettecken, T., Kraus, J., Meyer, C., Enders, A., Wirtz, A., Meitinger, T., et al. (2004). Tetrasomy 21pter–>q21.2 in a male infant without typical Down's syndrome dysmorphic features but moderate mental retardation. J. Med. Genet. *41*, e26.

Shaw, D.R. (2009). Searching the Mouse Genome Informatics (MGI) resources for information on mouse biology from genotype to phenotype. Curr. Protoc. Bioinformatics *Chapter 1*, Unit 1.7.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, http://dx.doi.org/10.2202/1544-6115.1027, Article 3.

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol. Syst. Biol. *8*, 608.

Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat. Biotechnol. *27*, 199–204.

Wang, X., Wei, X., Thyssen, B., Das, J., Lipkin, S.M., and Yu, H. (2012). Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat. Biotechnol. *30*, 159–164.

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). Web-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. *41*, W77–W83.

Watanabe, K., Sugawara, C., Ono, A., Fukuzumi, Y., Itakura, S., Yamazaki, M., Tashiro, H., Osoegawa, K., Soeda, E., and Nomura, T. (1998). Mapping of a novel human carbonyl reductase, CBR3, and ribosomal pseudogenes to human chromosome 21q22.2. Genomics *52*, 95–100.

Wisniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol. Cell Proteomics *13*, 3497–3506.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. *33*, W741–W748.

Zhou, X., Menche, J., Barabasi, A.L., and Sharma, A. (2014). Human symptoms-disease network. Nat. Commun. *5*, 4212.

STAR★METHODS

KEY RESOURCE TABLE

Reagent or Resource	Source	Identifier
Deposited Data		
Inweb_IM	Li et al., 2016	http://lagelab.org/resources
Gene expression data	Lockstone et al., 2007	GSE5390
Gene expression data	Mao et al., 2005	GSE1397
Microarray mRNA expression data	Stingele et al., 2012	GSE39768
Software and Algorithms		
R programming language		http://r-project.org
Other		
WebGestalt	Wang et al., 2013 Zhang et al., 2005	http://webgestalt.org
The Mouse Genome Informatics (MGI) resource	Shaw, 2009	http:/informatics.jax.org

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Søren Brunak (soren.brunak@cpr.ku.dk).

METHOD DETAILS

Protein Interaction Data

The InWeb_IM interactome data is a network of human protein-protein interactions (PPIs) based on experimental interaction data from humans and model organisms extracted from various protein interaction resources (Li et al., 2017). The network comprised 1,608,186 unique interactions covering 15,214 proteins encoded on chromosome 1-22, X and Y. In this analysis, the number of possible PPIs was restricted to exclude self-interactions as well as histone proteins and ubiquitin C (Horton et al., 2004) since these were two very highly connected proteins in the global network leaving 1,364,024 PPIs. We also removed tandem duplicated genes registered in the Duplicated Genes Database, so only one member was included from each family. All interactions in InWeb_IM are scored and benchmarked against a gold standard (Li et al., 2017) and PPIs with minimum confidence scores of 0.10 (N=251,401) are considered highly reliable. To assess the robustness of the pairwise comparison across chromosomes we used two additional subsets of InWeb_IM with scores $\geq$ 0.05 (N=396,959) and $\geq$ 0.15 (N=195,207).

Chromosome-Wise Connectivity Ratio

For each chromosome we calculated the connectivity ratio (CR), a measure that relates the number of *cis*-chromosomal interactions between proteins encoded on the same chromosome to the number of *trans*-chromosomal interactions that occur between proteins located on different chromosomes. CR is defined as:

$$CR = \frac{(n_C/N_C)}{(m_C/M_C)}$$

where n_C is the observed number of *cis*-chromosomal interactions on chromosome C normalized by the theoretically possible number of *cis*-chromosomal interactions on chromosome C given by $N_C = \frac{\{C\}(\{C\}-1)}{2}$, $\{C\}$ is the number of protein coding genes on chromosome C. The normalized number of *trans*-chromosomal interactions is likewise the number of observed *trans*-chromosomal interactions, m_C, normalized by the number of theoretically possible *trans*-chromosomal interactions, M_C, given by $\{C\} * \sum_{D,D \neq C} \{D\}$, where $\{D\}$ is the number of genes on any other chromosome than C. Self-interactions were excluded.

Global Similarity of Chromosomal PPI Profiles

For each pair of chromosomes in the human interactome, we counted the number of PPIs between pairs of genes located on both chromosomes. The observed number of interactions between all pairs of chromosomes was then normalized by the theoretically possible number of PPIs, which was defined as $\{C_A\}^*\{C_B\}$, where $\{C_A\}$ and $\{C_B\}$ represent the number of genes on chromosomes A and B. The resulting normalized counts were clustered using hierarchical clustering to compare connectivity ratios across the human chromosomes.

Highly Connected Hub Proteins

Based on the protein interaction data, we generated a list of all human proteins with minimum five first-order interaction partners found in InWeb_IM, we abbreviate these as hubs. The cut off for interaction partners was set for reliable expression correlation estimations between the central protein and its interaction partners.

Gene Expression Data

Gene expression data were obtained in previous studies from human adult brain tissue (dorsolateral prefrontal cortex) based on seven DS patients and eight healthy controls (Lockstone et al., 2007) and fetal brain tissue (cerebrum and cerebellum) based on seven trisomy 21 samples and seven aneuploidies (Mao et al., 2005). Both sets were analyzed using Affymetrix Human Genome U133A arrays. Microarray mRNA data were obtained in a previous study from human retinal pigment epithelial (RPE-1) cell lines based on three trisomy 21 aneuploidies and three diploids (Stingele et al., 2012), this set was analyzed by Agilent Whole Human Genome Microarray.

Data analysis was performed for each set separately using the R programming language (version 3.2.1). We used RMA (Irizarry et al., 2003) to normalize the raw expression signals for the gene expression data while background correction and normalization of the mRNA was done using backgroundCorrect and normalizeBetweenArrays from the R package limma (Smyth, 2004). Log2-fold-change values for all probes on the array was also done with limma (Smyth, 2004). Probes without detectable expression signals were removed and probe-wise fold-change values were summarized for each gene by averaging over all probe-wise fold-changes mapping to the gene.

Expression Correlation of Hub Proteins

Active hub proteins were identified based on correlated expression between the hub, *H*, and its interactors (Han et al., 2004), *I*, by using Pearson's correlation coefficient, PCC. The correlations were calculated across patients *p=1, 2, 3,.., n,* and done separately for the DS patients and controls to differentiate between hubs exclusively correlated in DS patients, in controls, or in both. Correlation coefficients for both groups were calculated as:

$$r_{I,H} = \frac{\sum_{p=1}^{n}(X_{I_p} - \overline{X}_I)(X_{H_p} - \overline{X}_H)}{(n-1)s_I s_H}$$

where $\overline{X}_I = \frac{\sum_{p=1}^{n}(X_{I_p})}{n}$, $\overline{X}_H = \frac{\sum_{p=1}^{n}(X_{H_p})}{n}$, $S_I = \sqrt{\sum_{p=1}^{n}\frac{(X_{I_p}-\overline{X}_I)^2}{(n-1)}}$ and $S_H = \sqrt{\sum_{p=1}^{n}\frac{(X_{H_p}-\overline{X}_H)^2}{(n-1)}}$. $S_I S_H$ is the product of the standard deviations (SD) of the expression data for the hub and its interactors. To define a single measurement of correlation for each hub, the average PCC (*AvgPCC*) over interactors, n_H, was calculated as:

$$AvgPCC = \frac{\sum_{I=1}^{n_H} r_{I,H}}{n_H}$$

The *AvgPCC* was calculated for DS patients and controls, respectively. To rank the hubs based on their correlations, *AvgPCCs* were normalized by transformation to z-scores:

$$z = \frac{AvgPCC - \mu(AvgPCC)}{\sigma(AvgPCC)}$$

We ranked the hubs based on their absolute z-scores, and hubs with |z| greater than 2 SD above mean (indicating they are the top 5% of most correlated hubs) in either DS patients, controls, or both groups were considered as active hub proteins. Three groups of hubs were identified based on the above threshold: 1) hubs gained in DS i.e. only correlated/active in DS; 2) hubs lost in DS, i.e. only correlated/active in controls; and 3) constitutively active hubs that were correlated/active in both (Table 1). All analyses were performed using the R programming language (version 3.2.1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Significance of the Chromosome-wise Connectivity Ratio (CR)

We assessed the significance of the CR in two ways. The first approach was to reassign proteins to new chromosomes randomly while maintaining the original interactions, and in the second approach we used random network rewiring where we shuffled all interactions while maintaining the degree for each node. Both approaches were done for 10,000 iterations and p values were recorded for the true CR being higher or lower than expected based on the fraction of iterations with an artificial CR higher or lower than the true value for the given chromosome. The randomization and subsequent significance calculation were both performed in the R programming language (version 3.2.1).

Disease Enrichment in Chromosome 21 Encoded Hubs

The hubs in either gained or loss of function in DS encoded by genes on chromosome 21 were investigated for enrichment in diseases association. Each group was tested against all the hubs encoded by genes on chromosome 21 using the web-based version of

WebGestalt (Wang et al., 2013; Zhang et al., 2005). A hypergeometric test was used as the statistical model, and p values for significance were adjusted using Benjamini-Hochberg correction.

Gene Enrichment

Enrichment of hubs encoded by genes at the different chromosomes, *C*, was calculated for each of the three groups, *i*, defined by the *AvgPCC*, as:

$$E_{i,C} = \frac{n_{H_C}/N_i}{m_{H_C}/M}$$

where $n_{H_{i,C}}$ is the number of hubs in group *i* that were encoded by a gene on chromosome *C*. This is normalized by the sum of all genes in group *i*, N_i. The enrichment was corrected with the background using m_{H_C}/M, where m_{H_C} is the number of hubs encoded by a gene on chromosome *C* in the entire dataset and *M* is the total number of hubs in the set.

The significance of the enrichment was calculated in the same manner as the first approach for the CR significance. The chromosome names were shuffled and reassigning to all the hubs and their average PCC, and group assignment was re-done and enrichment recalculated. This was done for 100,000 iterations and the p value for the true enrichment was calculated in the same manner as with the CR. The randomization and subsequent significance calculation were both performed in the R programming language (version 3.2.1).

Functional Sub-Classification of Correlated Hub Proteins

The functionality of the hubs in the three different groups (gained/lost in DS patients and constitutive active) was characterized by integrating information from the Mouse Genome Informatics (MGI) resource (Shaw, 2009). Uncorrelated hubs were used as a reference. Statistical analysis was performed using R programming language (version 3.2.1). Significance of phenotypic differences between groups was established based on a hypergeometric test.

ADDITIONAL RESOURCES

The protein interaction dataset, Inweb_IM, is available at (http://www.lagelab.org/resources/).